

ABSTRACT

A chip-multiprocessing system with scalable architecture, including on a single chip: a plurality of processor cores; a two-level cache hierarchy; an intra-chip switch; one or more memory controllers; a cache coherence protocol; one or more coherence protocol engines; and an interconnect subsystem. The two-level cache hierarchy includes first level and second level caches. In particular, the first level caches include a pair of instruction and data caches for, and private to, each processor core. The second level cache has a relaxed inclusion property, the second-level cache being logically shared by the plurality of processor cores. Each of the plurality of processor cores is capable of executing an instruction set of the ALPHA™ processing core. The scalable architecture of the chip-multiprocessing system is targeted at parallel commercial workloads. A showcase example of the chip-multiprocessing system, called the PIRAHNA™ system, is a highly integrated processing node with eight simpler ALPHA™ processor cores. A method for scalable chip-multiprocessing is also provided.

15